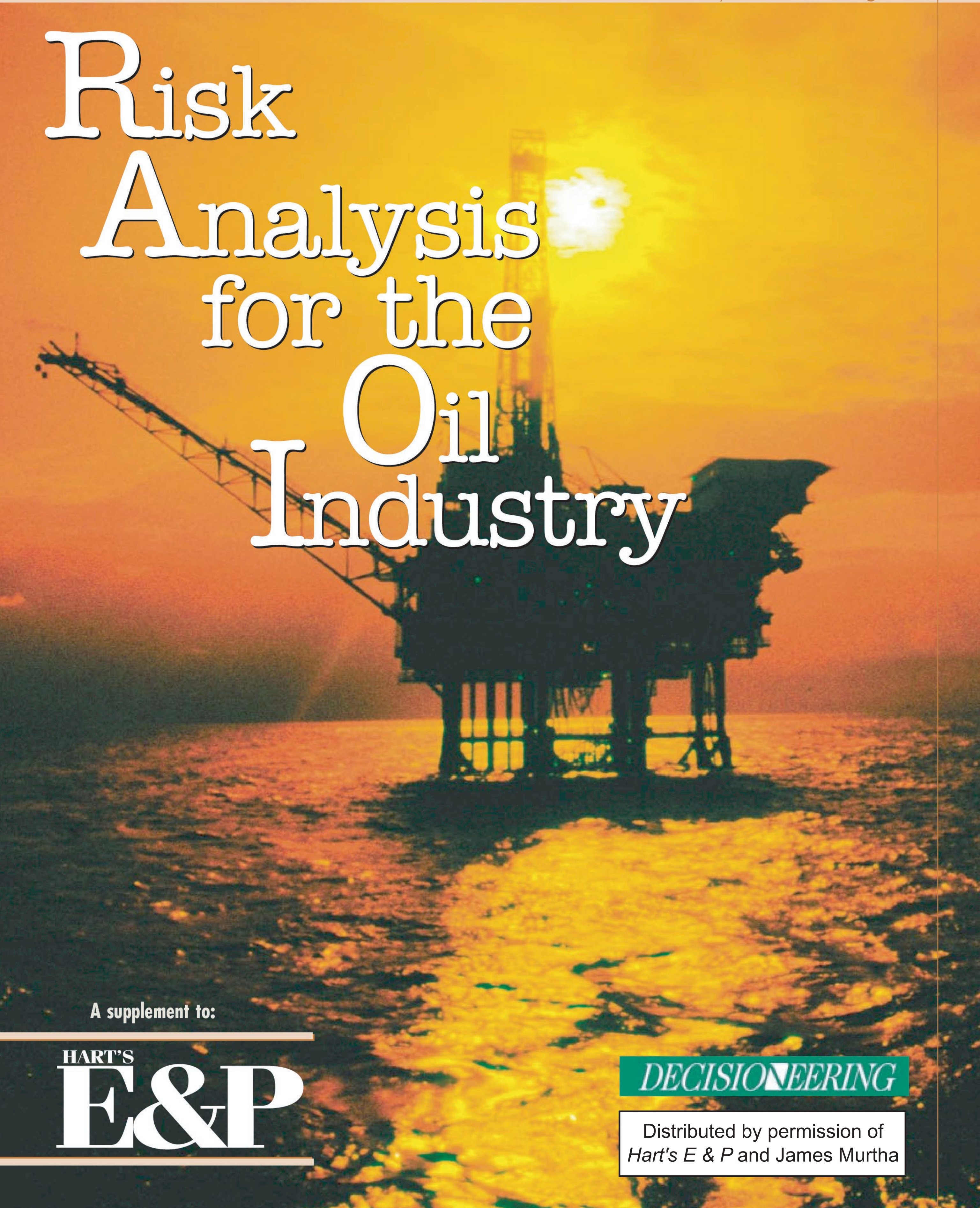
A Comprehensive
Study of Risk Management

# Risk Analysis for the Oil Industry



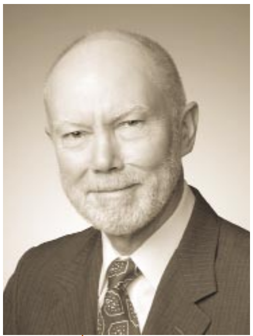
A supplement to:

HART'S
E&P

DECISIONEERING

Distributed by permission of
*Hart's E & P* and James Murtha

# Biography



Jim Murtha, a registered petroleum engineer, presents seminars and training courses and advises clients in building probabilistic models in risk analysis and decision making. He was elected to Distinguished Membership in SPE in 1999, received the 1998 SPE Award in Economics and Evaluation, and was 1996-97 SPE Distinguished Lecturer in Risk and Decision Analysis. Since 1992, more than 2,500 professionals have taken his classes. He has published Decisions Involving Uncertainty – An @RISK Tutorial for the Petroleum Industry. In 25 years of academic experience, he chaired a math department, taught petroleum engineering, served as academic dean, and co-authored two texts in mathematics and statistics. Jim has a Ph.D. in mathematics from the University of Wisconsin, an MS in petroleum and natural gas engineering from Penn State and a BS in mathematics from Marietta College. ◆

# Acknowledgements

*When I was a struggling assistant professor of mathematics, I yearned for more ideas, for we were expected to write technical papers and suggest wonderful projects to graduate students. Now I have no students and no one is counting my publications. But, the ideas have been coming. Indeed, I find myself, like anyone who teaches classes to professionals, constantly stumbling on notions worth exploring.*

*The articles herein were generated during a few years and written mostly in about 6 months. A couple of related papers found their way into SPE meetings this year.*

*I thank the hundreds of people who listened and challenged and suggested during classes.*

*I owe a lot to Susan Peterson, John Trahan and Red White, friends with whom I argue and bounce ideas around from time to time.*

*Most of all, these articles benefited by the careful reading of one person, Wilton Adams, who has often assisted Susan and me in risk analysis classes. During the past year, he has been especially helpful in reviewing every word of the papers I wrote for SPE and for this publication. Among his talents are a well tuned ear and high standards for clarity. I wish to thank him for his generosity.*

*He also plays a mean keyboard, sings a good song and is a collaborator in a certain periodic culinary activity.*

*You should be so lucky.* ◆

# Table of Contents

# Central Limit Theorem – Polls and Holes

What do exit surveys of voters in presidential elections have in common with porosities calculated from logs of several wells penetrating a geological structure? The answer is that in both cases, the data can be used to estimate an average value for a larger population.

At the risk of reviving a bitter debate, suppose that a carefully selected group of 900 voters is surveyed as they leave their polling booths. If the voters surveyed a) are representative of the population of a whole and b) tell the truth, then the ratio

**r = (Number of voters for Candidate A in survey)/(number of voters surveyed)**

should be a good estimate of the ratio.

**R = (Number of voters for Candidate A in population)/(number of voters in population)**

Moreover, by doing some algebra, the statistician analyzing the survey data can provide a margin of error for how close r is to R. In other words, you are pretty confident that

**(r - margin of error) < R < (r + margin of error)**.

The margin of error formula depends on three things:

- the level of confidence of the error ("I am 90% or 95 % or 99% confident of this");
- the number of voters who chose candidate A; and
- the number of voters surveyed.

To end this little diversion, here is the approximate formula for the margin of error in a close race (where R is roughly 45% to 55%), and we are satisfied with a 95% confidence level (the most common confidence level used by professional pollsters).

**Margin of error = 1/sqrt (N) approx.**

**Where N = sample size, the number of voters polled.**

Thus when **N=900**, the margin of error would be about 1/30 or 3%.

Thus, if 54% of the voters surveyed said they voted for candidate A, then there is about a 95% chance that A will get between 51% and 57% of the full vote.

This method of analysis is a direct consequence of the Central Limit Theorem, one of the most significant results in mathematical statistics.

Suppose you have a distribution X, with mean $\mu$ and standard deviation $\sigma$. X can be virtually any shape. From X, we sample n values and calculate their mean. Then we take another sample of size n from X and calculate its mean. We continue this process, obtaining a large number of samples and build a histogram of the sample means. This histogram will gradually take a shape of a smooth curve. The amazing fact is that that curve, the limit of the sample means, satisfies these conditions:

- the sample means are approximately normally distributed;
- has mean equal $\mu$, but;
- has standard deviation of approximately $\sigma\sqrt{n}$ (the mean standard error).

The key here is that the mean standard error gets small as n (the sample size) gets large.

We need another fact about normal distributions, namely that 68% of the values in a normal distribution lie within one standard deviation of its mean, and 95% and 99.7% of the values lie within two and three standard deviations, respectively.

In our case, X is a binomial variable with exactly two possible values, 0 and 1, where the probability of 1 is the percentage, p, of people among all voters who voted for A and the probability of 0 is 1-p. The mean value of this distribution is p. The standard deviation is $\sqrt{p(1-p)}$.

**Here's the point: the Central Limit Theorem guarantees these distributions of average properties – net pay, porosity and saturation – will tend to be normal.**

# This would be a good time to ask that important question, "Why bother?"

| Structure | Area | Well | Net pay | Porosity | Sw |
|---|---|---|---|---|---|
| 1S | 5010 | 1S1 | 34 | 0.12 | 0.31 |
| 1S | 5010 | 1S2 | 56 | 0.07 | 0.20 |
| 1S | 5010 | 1S3 | 74 | 0.14 | 0.38 |
| 1S | 5010 | 1S4 | 47 | 0.09 | 0.30 |
| 1S | 5010 | 1S5 | 53 | 0.18 | 0.26 |
| 1S | 5010 | 1S6 | 29 | 0.08 | 0.29 |
| 1S | 5010 | 1S7 | 84 | 0.15 | 0.37 |
| 1S | 5010 | 1S8 | 40 | 0.14 | 0.25 |
| 2S | 2600 | 2S1 | 35 | 0.12 | 0.33 |
| 2S | 2600 | 2S2 | 42 | 0.07 | 0.38 |
| 2S | 2600 | 2S3 | 37 | 0.09 | 0.41 |
| 2S | 2600 | 2S4 | 61 | 0.13 | 0.26 |
| 2S | 2600 | 2S5 | 27 | 0.10 | 0.33 |
| 2S | 2600 | 2S6 | 48 | 0.08 | 0.35 |
| 3S | 4300 | 3S1 | 68 | 0.12 | 0.26 |
| 3S | 4300 | 3S2 | 94 | 0.16 | 0.20 |
| 3S | 4300 | 3S3 | 35 | 0.07 | 0.37 |
| 3S | 4300 | 3S4 | 47 | 0.09 | 0.30 |
| 3S | 4300 | 3S5 | 68 | 0.14 | 0.32 |
| 3S | 4300 | 3S6 | 75 | 0.10 | 0.28 |
| 3S | 4300 | 3S7 | 67 | 0.12 | 0.25 |
| 3S | 4300 | 3S8 | 48 | 0.13 | 0.23 |
| 3S | 4300 | 3S9 | 69 | 0.09 | 0.30 |
| 3S | 4300 | 3S10 | 88 | 0.15 | 0.23 |

*Table 1. Well database*

| Structure | Area | h_ave | Por_ave | Sw_ave |
|---|---|---|---|---|
| 1S | 5010 | 52 | 0.12 | 0.30 |
| 2S | 2600 | 42 | 0.10 | 0.34 |
| 3S | 4300 | 66 | 0.12 | 0.27 |
| 4S | … | … | … | … |

*Table 2. Structure database*

Nevertheless, in many applications, as we shall see later, X is a continuous variable, with a shape of a normal, lognormal or a triangular distribution, for example.

Thus, our sample's mean (the percentage of the people in our sample who voted for A) may not be exactly the mean of the normal distribution, but we can be 95% confident that it is within two values of the mean standard error of the true mean.

What does this have to do with porosity?

## Application: average net pay, average porosity, average saturations

Consider the following typical volumetric formula:

$G = 43{,}560Ah(1-Sw)\varphi/Bg{*}E$

Where A = Area
h = Net pay
$\varphi$ = Porosity
Sw = Water saturation
Bg = Gas formation volume factor
E = Recovery efficiency

In this formula, there is one component that identifies the prospect, A, while the other factors essentially modify this component. The variable h, for example, should represent the average net pay over the area A. Similarly, $\varphi$ represents the average porosity for the specified area, and Sw should represent average water saturation.

Why is this? Because, even though there may be a large range of net pay values throughout a given area, A, we are interested only in that average net pay which, when multiplied by A, yields the (net sand) volume of the prospect.

Consider a database of wells in some play, as shown in Table 1, grouped into structures. Structure 1 has eight wells, structure 2 has six wells, and so on.

We need to take this a step further. Suppose we construct a net sand isopach map for structure 1, and then calculate the average net sand for the structure, by integrating and dividing by the base area. Then we do the same thing for porosity as well as for water saturation. This yields a new database, Table 2, which becomes the basis for a Monte Carlo simulation. In the simulation, when we sample a value for A, we then sample values for average net pay, average porosity and average saturation for that area, which gives one realization of gas in place.

Although the suggested process of obtaining the averages involved integration, one could use the numerical average of the data to approximate the integrated average. Even so, since we often try to drill wells in the better locations, it would still be useful to sketch a map with the well locations and get some idea whether the well average would be biased compared with the averages over the entire structure. If we put a grid over the area and place a well location in each grid block, then the two notions of average essentially coincide.

## These distributions of averages are normal

Here's the point: the Central Limit Theorem guarantees these distributions of average properties – net pay, porosity and saturation – will tend to be normal. Another consequence of the theorem is that these distributions of averages are relatively narrow, for example, they are less dispersed than the full distributions of net pays or porosities or saturations from the wells, which might have been lognormal or some other shape. The correct distributions for Monte Carlo analysis, however, are the narrower, normal-shaped ones.

One objection to the above argument (that we should be using narrower, normal distributions) is that often we do not have ample information to estimate the average porosity or average saturation. This is true. Nonetheless, one could argue that it still might be possible to imagine what kind of range of porosities might exist from the best to the worst portions of the structure.

## How much difference will this make?

This would be a good time to ask that important question, "Why bother?" What if this is correct: that we should be using narrower and more symmetric shaped distributions for several of the factors in the volumetric formula? Does it matter in the final estimate for reserves or hydrocarbon in place? How much difference could we expect?

Let us consider porosity. When we examine the database of porosities from all wells (for example, the average log porosity for the completion interval or from the layer interval in a multilayer system) as in Table 1, there are two possible extreme situations. At one extreme, it is possible that each structure exhibits a wide range of quality from fairway to flanks, but the average porosities for the various structures always fall between 10% and 14%. In this case, the range of all porosities could easily be several times as broad as the average porosities. That is, there are similarities among structures, but not much homogeneity within a structure.

On the other hand, it is possible that each structure is relatively homogeneous, but the different structures are quite dissimilar in quality, with average porosities

**Even the main factor (area or volume) can be skewed left.**

ranging from 5 % to 20%. In this case, the two distributions would be closer together. Note, however, that the average porosities will always have a narrower distribution than the complete set of porosities.

Perhaps the contrast is even easier to see with net pays. Imagine a play where each drainage area tends to be relatively uniform thick, which might be the case for a faulted system. Thus, the average h for a structure is essentially the same as any well thickness within the structure. Then the two distributions would be similar. By contrast, imagine a play where each structure has sharp relief, with wells in the interior having several times the net sand as wells near the pinch outs. Although the various structures could have a fairly wide distribution of average thicknesses, the full distribution of h for all wells could easily be several times as broad. The distribution for A could easily be lognormal if the drainage areas were natural. In a faulted system, however, where the drainage areas were defined by faults, the distribution need not be lognormal. The

To help your imagination, we do have ample information from many mature fields where material balance could provide estimates. We also have extensive databases with plenty of information, from which some range of average values could be calculated and compared with the broader ranges of well data.

Always remember that, like all else in Monte Carlo simulation, you must be prepared to justify every one of your realizations, for example, combinations of parameters. Just as we must guard against unlikely combinations of input parameters by incorporating correlations in some models, we should ask ourselves if a given area or volume could conceivably have such an extreme value for average porosity or average saturation. If so, then there must be even more extreme values at certain points within the structure to produce those averages (unless, of course, the structure is uniformly endowed with that property, in which case, I would be skeptical of the wide disparity from one structure to the next.)

**Additional complications arise because of uncertainty about the range of driving mechanisms – will there be a water drive? Will gas injection or water injection be effective?**

right way to judge whether the type of distribution matters for an input variable is to compare what happens to the output of the simulation when one type is substituted for another.

### What about the output of the simulation, OOIP?

Regardless of the shapes of the inputs to a volumetric model – be they skewed right, or left, or symmetric the – output will still be skewed right, thus approximately lognormal. In fact, as is well known, the Central Limit Theorem guarantees this. The argument is straightforward: the log of a product (of distributions) is a sum of the logs (of distributions), which tends to be normal. Thus the product, whose log is normal, satisfies the definition of a lognormal distribution.

### Will the Area distribution always be lognormal?

The traditional manner of describing area and treating it as a lognormal distribution is based on prospects in a play. If we were to select at random some structure in a play, then the appropriate distribution likely would be a lognormal. Sometimes, however, not even the Area parameter should be modeled by a lognormal distribution. Why? Suppose a particular prospect is identified from 3-D seismic. We have seen situations where the base case value of area or volume is regarded as a mode (most likely). When asked to reprocess and or reinterpret the data and provide relatively extreme upside (say P95) and downside (say P5) areas or volumes, the results are skewed left: there is more departure from the mode toward the downside than the upside. Because the conventional lognormal distribution is only skewed right, we must select another distribution type, such as the triangular, beta, or gamma distribution.

### Variations of the volumetric formula

Among the numerous variations of the volumetric formulas, there is usually only one factor that serves the role of Area in the above argument. For instance, another common formula is:

$$OOIP = 7{,}758 Vb(NTG)\phi So/Bo$$

Vb = Bulk rock volume

NTG = Net to gross ratio

Here, Vb would be the dominant factor, which could be skewed right and modeled by a lognormal distribution, while the factors NTG, $\phi$, So and Bo would tend to be normally distributed, since they represent average properties over the bulk volume.

### Recovery factors

Recovery factors, which convert hydrocarbon in place to reserves or recoverable hydrocarbon, are also average values over the hydrocarbon pore volume. The recovery efficiency may vary over the structure, but when we multiply the OOIP by a number to get recoverable oil, the assumption is that this value is an average over the OOIP volume. As such, they too would often be normally distributed. Additional complications arise, however, because of uncertainty about the range of driving mechanisms: will there be a water drive? Will gas injection or water injection be effective? Some people model these aspects of uncertainty with discrete variables.

### Summary

The Central Limit Theorem suggests that most of the factors in a volumetric formula for hydrocarbons in place will tend to have symmetric distributions and can be modeled as normal random variables. The main factor (area or volume) can be skewed left or right. Regardless of the shapes of these input distributions, the outputs of volumetric formulas, oil and gas in place and reserves, tend to be skewed right or approximately lognormal.

Because the conventional wisdom is to use lognormal distributions for all of the inputs, the above argument may be controversial for the time being. The jury is still out. We could take a poll and see what users believe. Oh yes, then we could use the Central Limit Theorem to analyze the sample and predict the overall opinion.

What goes around comes around.

Stay tuned for other applications of the Central Limit Theorem. ◆